HHS Public Access
Author manuscript
Qual Life Res. Author manuscript; available in PMC 2016 April 15.

Published in final edited form as:
Qual Life Res. 2016 April ; 25(4): 823–833. doi:10.1007/s11136-015-1143-z.

Linkage between the PROMIS® Pediatric and Adult Emotional Distress Measures

Bryce B. Reeve[1,2], **David Thissen**[3], **Darren A. DeWalt**[4], **I-Chan Huang**[5,6], **Yang Liu**[3], **Brooke Magnus**[3], **Hally Quinn**[3], **Heather E. Gross**[2], **Pamela A. Kisala**[7], **Pensheng Ni**[8], **Stephen Haley**[8], **MJ Mulcahey**[9], **Susie Charlifue**[10], **Robin Hanks**[11], **Mary Slavin**[8], **Alan Jette**[8], and **David S. Tulsky**[7]

[1]Department of Health Policy and Management, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

[2]Cecil G. Sheps Center for Health Services Research, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

[3]Department of Psychology, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

[4]Division of General Medicine and Clinical Epidemiology, Cecil G. Sheps Center for Health Services Research, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

[5]Department of Health Outcomes and Policy, College of Medicine, University of Florida, Gainesville, FL, USA

[6]Department of Epidemiology and Cancer Control, St. Jude Children's Research Hospital, Memphis, TN, USA

[7]Center for Assessment Research and Translation, College of Health Sciences, Department of Physical Therapy, University of Delaware, Newark, DE, USA

[8]Boston University School of Public Health, Health and Disability Research Institute, Boston, MA, USA

[9]Department of Occupational Therapy, School of Health Professions, Thomas Jefferson University, Philadelphia, PA, USA

[10]Craig Hospital, Englewood, Colorado, USA

[11]Rehabilitation Institute of Michigan, Detroit, USA

Correspondence concerning this article should be addressed to: Bryce B. Reeve, PhD, University of North Carolina, 1101-D McGavran-Greenberg Building, 135 Dauer Drive, CB 7411, Chapel Hill, NC 27599-7411, Phone: 919-843-8793, Fax: 919-843-6308, bbreeve@email.UNC.edu.

Compliance with Ethical Standards

Conflict of Interest: Drs. DeWalt and Tulsky were unpaid members of the Board of Directors for the PROMIS Health Organization (PHO) during the conduct of this study. Drs. Reeve and Tulsky were unpaid members of the Board of Directors for the PHO during the preparation of this manuscript. The remaining authors have no financial relationships or conflicts of interest relevant to this study to disclose.

Research involving Human Participants: All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed Consent: Informed consent was obtained from all individual participants included in the study.

Introduction

The goal of the Patient-Reported Outcomes Measurement Information System® (PROMIS®) initiative was to design standardized patient-reported outcome (PRO) measures that can be used in research studies that include patients with various diseases and the general population. Pediatric and adult PROMIS measures were developed in parallel and, in theory, measure many of the same PRO domains; however, they were built differently, with the target age population in mind to create measures that are age-appropriate in content and language. This includes cognitive interviews with participants from the target age groups to refine the measures [1; 2]. The purpose of this study is to create a linking algorithm between the PROMIS pediatric and adult measures for the emotional distress domains to facilitate PRO measurement across the lifespan. Having separate pediatric and adult versions of a PRO measure presents challenges when one wants to compare or combine data from studies that include both pediatric and adult participants, or when a longitudinal study begins data collection in childhood and continues into adulthood. Linking pediatric and adult PRO measures to provide comparable scores gives researchers a powerful tool in the study of diseases that affect both children and adults and childhood diseases with sequelae in adulthood.

The linking of PRO measures has not received extensive attention in the health outcomes literature. Dorans provides an overview of scale linking methods [3], and the PROMIS pediatric and adult scales use the "item banking" concept [4; 5], which uses item response theory (IRT) calibration method to link scores on alternate short forms and those from computerized adaptive tests (CATs). Among PRO measures, researchers have linked alternate forms of similar measures of functional health status activity [6], physical functioning [7], self-regulation [8], and depression [9]. The linking of pediatric and adult PRO measures of physical functioning has been conducted, but was limited to a specific population of individuals with spinal cord injury [10]. Others have attempted linkage of disparate PRO measures, such as the Medical Outcomes Study Short Form-36 with the Louisiana State University Health Status Instruments physical functioning scales [11; 12] and the European Organization for Research and Treatment of Cancer Core Questionnaire with the Functional Assessment for Cancer Therapy – General physical, emotional, and role/functional subscales [13], and a set of eleven depression scales [14]. These attempts at linking demonstrate the power of IRT, but also raise questions about when and what to link in measures of health outcomes. A cautionary note is that PRO scales may have the same name, but this does not necessarily mean they measure the same construct [15]. The extent to which separately developed scales measure the same construct is an empirical question. The answer to this question is an important prerequisite for determining the feasibility of linking the PROMIS pediatric and adult measures.

This study evaluates the viability of linking the PROMIS pediatric and adult measures and then applies a relatively new statistical procedure called *calibrated projection* [16] that uses IRT to link the two measures. To enhance the confidence in the generalizability of this linking methodology, we carry out the analyses separately in two different populations and compare the results to look at the stability of the linking. One sample included adolescents and young adults (ages 14–20 years) with "special health care needs" who have or are at increased risk for a chronic physical, developmental, behavioral, or emotional condition and who also require healthcare or related services beyond what patients generally require. The second sample included adolescents and young adults (ages 14–24 years) who have “physical or cognitive disabilities”, including those diagnosed with spinal cord injury, traumatic brain injury, or cerebral palsy. This article focuses on the linkage of the emotional distress domains of the PROMIS measures (i.e, Depressive Symptoms, Anxiety, and Anger) [17–19] while a future study will focus on the linkage of the physical health domains (i.e., physical function, pain, fatigue).

Methods

Research Participants and Data Collection

Sample 1: Individuals with “Special Health Care Needs”—Adolescents and young adults with special health care needs (SHCN) [20] were recruited because they represent a diverse set of illnesses that affects all domains of health-related quality of life measured by PROMIS. Inclusion criteria included the individual must have SHCN; be between 14–20 years of age; be able to read, write and speak English; and have access to a computer with an internet connection.

The sample of individuals with SHCN was collected from two sources: public health insurance programs (Medicaid and Children’s Health Insurance Program [CHIP] in Florida) and the Opinions for Good (Op4G) panel. The status of SHCN of individuals was defined by the Clinical Risk Groups (CRGs) [21] in the Medicaid/CHIP sample and by the Special Health Care Needs Screener [22] in the Op4G sample. We planned to collect data from at least 800 individuals (400 adolescents14–17 years old and 400 young adults 18–20 years old).

From the Medicaid/CHIP databases, we randomly selected 20,000 individuals with SHCN, and 17,435 of those had telephone numbers for contact. After a maximum of 5 redials, we were unable to contact 11,806 individuals. For those who were contacted, 1,052 individuals refused to participate and 2,884 were excluded (either not English speaking or no computer and/or internet at home). For the 1,194 who verbally agreed to participate (666 adolescents and 528 young adults), invitation e-mails were sent to parents that included the survey link, username, and password. A $20 gift card was provided for individuals who completed the survey. Data collection was conducted between April 1st, 2012 and May 31st, 2013.

The remaining sample was collected from Op4G, a survey research company that partners with non-profits organizations so that panelists can earn money for charities by completing surveys. Op4G identified 320 adolescents (14–17 years old) and 320 young adults (18–20 years old) from the database and e-mailed survey invitations to parents with eligible adolescents and young adults. Participants earned $25, of which they could elect to donate between 25–100% to the non-profit agency of their choice. Data were collected between August 1st and September 30th, 2013.

For both Medicaid/CHIP and Op4G participants, consent forms were placed at the front of the survey to provide full disclosure of the study. Informed consent was obtained for young adults and parents of adolescents, and assent was obtained for adolescents. IRB approval was obtained at the University of North Carolina, the coordinating center, and University of Florida, who conducted and oversaw data collection.

Sample 2: Individuals with Physical or Cognitive Disabilities—Adolescents and young adults with traumatic spinal cord injury (SCI), traumatic brain injury (TBI), and/or cerebral palsy (CP) were recruited to ensure that PROMIS linkages are appropriate for adolescents and young adults with physical and/or cognitive disabilities. For all participants, a diagnosis of SCI, TBI, or CP was confirmed by medical record review. Individuals with CP must have received their diagnosis between the ages of 2–15 years to be eligible. Individuals with uncomplicated Mild TBI (i.e., Glasgow Coma Scale 13–15 with no positive neuroimaging findings) or non-traumatic SCI were excluded. Other inclusion criteria were the ability to read and understand English and respond to self-report scales by speaking, using a communication board, or gesturing. Participants ages 18–24 and parents of participants ages 14–17 provided informed consent, and participants ages 14–17 also provided assent. Potential participants were identified by study site coordinators at the collaborating sites: University of Michigan, Boston University, Craig Hospital, Rehabilitation Hospital of Michigan, and the Shriners Hospitals for Children (Philadelphia and Chicago locations). The University of Michigan served as the coordinating center, and IRB approval was obtained at all participating institutions.

All data were collected between June 1, 2011 and April 10, 2012 using the Assessment Center^{SM} data collection platform [23]. Because individuals with the most severe disabilities might be excluded from the study due to difficulties completing the scales independently using Assessment Center, the second sample was collected entirely using an interview mode of administration. Trained interviewers administered the items in interview format (either in person or through a telephone interview) to facilitate participation by individuals with physical limitations in using a computer or cognitive difficulties that might affect the sustained attention needed to complete items independently. Because response formats change across items, printed response cards were placed in front of participants to ensure that the correct response set was used for each item. When telephone interviews were conducted, participants were sent the response cards and the interviewer instructed them which card to use. Interviewers entered the responses into Assessment Center. Adolescents and young adults received $40 for study participation.

Measures

All participants completed a demographic questionnaire capturing the respondent's age, sex, race, ethnicity, and education level. Participants with SHCN completed additional items related to presence of any self-reported health conditions, while participants with disabilities completed additional items related to country of birth, language(s) spoken at home, methods of mobility, and secondary medical complications (e.g., neurogenic bowel/bladder). Then, all participants completed pediatric PROMIS short forms and corresponding adult PROMIS short forms for the following domains: Physical Function, Pain, Fatigue, Social Health, Depression, Anxiety, and Anger; see Table 1 for a detailed list.

Order of administration of the pediatric and adult PROMIS measures was randomized. Scores are on the PROMIS *T*-score metric with mean 50 and standard deviation of 10 in the original PROMIS calibration sample [24]. Higher scores on the PROMIS Depression, Fatigue, Anxiety, Pain, and Anger measures are associated with more severe symptoms, while higher scores on the physical functioning and social health measures are associated with better functioning.

Statistical Methods

All data analyses were done separately within each of the two samples, using each to cross-check the results from the other as appropriate, either by comparing the values of descriptive statistics, or by explicit double cross-validation at the completion of the modeling exercise.

We computed IRT response-pattern scale scores for each participant on the pediatric and adult measures, and tabulated the means, standard deviations, and the correlations between the pediatric and adult scores. While the correlations between the scores are attenuated by measurement errors, they provide some evidence about the type of linkage that might be useful. Dorans [25; 26] suggested a minimum correlation of 0.866 between scores to obtain a component of the standard error due to unidimensional linking that is less than 0.5.

To evaluate the viability of linking between the pediatric and adult scales, we examine the population-invariance of (potential) linking using the Root Expected Mean Square Difference (REMSD) statistic of Dorans and Holland [27] for males vs. females and for adolescents vs. young adults. If the pediatric and adult scales measure very different constructs, the differences between score means for males and females or between adolescents and young adults may vary between them. These grouping variables are well suited to checking the viability of linking, because the REMSD essentially compares the sex-related differences in the scores. If two scales measure the same construct, sex- or age-related differences should be approximately the same using results from either scale. Previous results suggest that a suitable criterion for “about the same” is 0.05–0.08 in the REMSD metric [27]. If the values of the REMSD statistic are sufficiently small, this evidence of invariance across subgroups justifies proceeding with the linkage. Conversely, if the REMSD values are too large, we would forego linkage, because it would not be the same for all respondents.

We also fit confirmatory two-dimensional IRT models to the data for each domain, with one factor for the pediatric items and the other factor for the adult items. All IRT models are fitted using the software IRTPRO [28], using the graded item response model [29; 30] that is used for all PROMIS measures. These models are used to estimate the disattenuated correlation between the pediatric and adult latent variables. In the unlikely event that the latent variable correlations are nearly 1.0, we could use the unidimensional models as the basis of linkage by score alignment between the pediatric and adult measures; such linkage would be symmetrical, because the two scales would measure the same construct.

However, we expected that the pediatric and adult latent variables would be highly, but not perfectly, correlated. In that eventuality, we proceed with linkage using projection of the pediatric scores onto the adult scale, and vice versa, using the linear approximation to calibrated projection [16]. When two scales being linked do not measure the same construct, linkage is not symmetrical. Instead, it is a matter of predicting one score from item responses on the other scale, and the expected value of the score is computed using a regression function.

Calibrated projection [31] uses IRT to link two measures, without considering the scores on the predictor scale to be fixed, and without the demand of conventional calibration (alignment) that the two are measures of the same construct. In calibrated projection, a two-dimensional IRT model is fitted to the item responses from the two measures: in the most commonly-used notation, θ_1 represents the underlying construct measured by the first scale, and θ_2 represents the underlying construct measured by the second scale. The correlation between θ_1 and θ_2 is estimated.

The linear approximation to calibrated projection method [16] uses a linear regression model to compute projected scores, and a two-component approximation to the error variance of the projected score that combines the error variance in the observed θ_1 estimate with the projection error variance, the latter due to the fact that the two constructs are not perfectly related. The parameters of the regression model are computed using estimates of the means and covariance matrix of the latent variables in a two-dimensional IRT model fitted to the current data. In the case of this application, the banked unidimensional item parameters for the PROMIS pediatric and adult scales are used in the item part of the model that links all of the results back to the original reference populations. The details of this procedure have been described elsewhere [16].

Using the linear approximation to calibrated projection, the projected values of the IRT Expected a Posteriori scale score $\widehat{\mathrm{EAP}}[\theta_2]$ and its posterior standard deviation $\widehat{\mathrm{SD}}[\theta_2]$ are computed as

$$\widehat{\mathrm{EAP}}[\theta_2]=\beta_0+\beta_1\mathrm{EAP}[\theta_1] \quad (1)$$

and

$$\widehat{\mathrm{SD}}[\theta_2]=\sqrt{\beta_1^2\mathrm{SD}^2[\theta_1]+\mathrm{MSE}} \quad (2)$$

in which EAP[θ_1] and SD[θ_1] are the values of the scale score and the posterior standard deviation on the scale on which the respondent has been measured, β_0 and β_1 are the intercept and slope of a regression equation computed using the IRT estimate of the population means and covariance matrix of the two latent variables, and MSE is the mean squared error of the regression equation similarly computed. We note in passing that, in the unlikely event that a unidimensional IRT can be used for both the pediatric and adult scales simultaneously, equations (1) and (2) would also be used to compute estimates of scores on one scale from scores on the other: In that case, the values of β_1 and β_0 would be the transformation constants usually referred to as A and B [32], and the value of MSE would be zero.

Thus, all that is required is a tabulation of the values of β_0, β_1, and MSE for each domain, for each "direction": (1) when the pediatric scale score is known (θ_1) and the adult scale score is projected (θ_2), and (2) when the adult scale score is known (θ_1) and the pediatric scale score is projected (θ_2). We follow the recommendation that the accuracy of the linear approximation be checked against those obtained with calibrated projection for summed score conversion tables, which span the range of θ_1 and θ_2 [16].

The quality of the linkage can also be evaluated using statistics describing confidence interval coverage: If the linking is successful, 68% of the time the difference between the projected scale scores $\widehat{\mathrm{EAP}}[\theta_2]$ and the actual observed scale scores EAP[θ_2] (which we have, for the data from which the regression coefficients are derived as well as the cross-validation sample), should be within 1 $\widehat{\mathrm{SD}}[\theta_2]$; 95% of the time the difference should be within 2 $\widehat{\mathrm{SD}}[\theta_2]$s. We tabulate these proportions for each of the two samples' linear projections, and for each sample used as cross-validation for regression models with parameters estimated in the other sample. Finally, if the regression models obtained independently from the two samples are similar, we compute these proportions using the average of the two regression models for each domain and direction of projection, to provide evidence that the combined model performs as expected.

Results

The sample of 874 individuals with SHCN was diverse in respect to demographics including 53.2% female, 38.2% Hispanic, and 20.8% black, and in respect to health conditions, with greater than 10% of the sample experiencing hypertension, attention deficit hyperactivity disorder (ADHD), mental health condition, chronic pain, asthma, overweight, diabetes, and allergies (see Table 2). The sample of 641 individuals with disabilities included 36.8% females and 10.0% blacks; 30.7% had cerebral palsy, 37.8% had spinal cord injury, and 31.5% had a traumatic brain injury.

Table 3 shows the means and standard deviations of the scale scores (EAP[θ], on the standard PROMIS *T*-score metric) for the pediatric and adult scales, and the correlation coefficients (r) between the pediatric and adult scores. Individuals with SHCN exhibited higher levels of depressive symptoms, anxiety, and anger, and larger variation, than the sample of individuals with disabilities. The correlations between the pediatric and adult scale scores fell in the range 0.77 – 0.86; falling below the value of 0.866 that Dorans [25] recommended as a lower bound for unidimensional linking, suggesting that any linkage will take the form of projection.

Table 4 displays REMSD statistics for the sex difference and the difference between adolescents and young adults for the three emotional distress scales in the two samples. All of the REMSD values are less than the 0.05 – 0.08 range suggested for useful linkage, except for the sex difference for Depressive Symptoms scores in the sample of SHCN individuals; that high value is not replicated in the sample of individuals with disabilities, which makes it difficult to conclude that the sex difference varies reliably between the pediatric and adult scales. So, the REMSD statistics present no obstacle to linking.

The estimated latent-variable correlations (ρ) between the pediatric and adult constructs in Table 4 are in the range 0.87 – 0.94. While those correlations are high, their standard errors are all 0.01 to two decimal places, so they differ from 1.0. That means that linking using simultaneous calibration with a unidimensional IRT model is not appropriate. So we proceeded with linkage using calibrated projection, separately for each domain within each sample, using the linear approximation methods described by Thissen, Liu, Magnus, and Quinn [16].

Table 5 contains the values of the regression coefficients β_0 (intercept) and β_1 (slope), their 95% confidence intervals, and MSE that are used in equations 1 and 2 to compute the scale scores by calibrated projection from θ_1 to θ_2. For all but one of the twelve comparisons of coefficients across samples, the confidence intervals overlap. More importantly, the regression lines themselves are very similar between samples: None differ by more than 3 *T*-score points (0.3 standard deviation units) anywhere over the entire range of the *x*-axis variables. So the regression functions and values of MSE obtained from the two samples are extremely similar; the average values across the two samples that will be used subsequently are also shown.

We also computed the summed-score conversion tables for projected IRT scale scores and the posterior standard deviations to be used as their standard errors using both the linear approximation and calibrated projection to check the accuracy of the approximation. The estimates of the projected scores themselves were essentially identical, and the approximate standard deviations were 0.9 – 1.3 times the calibrated projection values, consonant with the results reported previously [16].

Table 6 tabulates the proportions of values of observed EAP[θ_2] that are within ±1 SD and ±2 SD of the values obtained using the linear approximation for each domain, within each sample, using all three sets of regression coefficients in Table 5. The values (proportions) for ±1 SD should be approximately 0.68, and actually fall in the range 0.58 – 0.76; those for ±2 SD should be about 0.95, and actually fall in the range 0.88 – 0.96. The values in Table 6 for which the parameters were obtained with the same sample are a check of model fit; the values for parameters obtained with the other sample are double cross-validation. The values in the lower third of the table show that the average values of the parameters obtained from the two separate samples perform very well: The values for ±1 SD fall in the range 0.67 – 0.76; those for ±2 SD are in the range 0.92 – 0.96.

Discussion

Each of the three PROMIS pediatric and adult emotional distress scales (Depressive Symptoms, Anxiety, and Anger) measure closely related but not identical constructs. We have used the linear approximation to IRT-based calibrated projection to provide the parameters for regression models that may be used to project scores either from responses to the PROMIS pediatric measure onto the PROMIS adult measure, or from responses to the PROMIS adult measure onto the PROMIS pediatric measure.

While there are uses for linked scores of this kind, it should be emphasized that higher-quality measurement is always obtained by using the intended measure instead of projecting results from some other measure. The reason for this is clear in equation (2): the projected posterior standard deviation, which is reported as the standard error of the projected score, is always larger (by a factor of the regression MSE) than the standard error would be if there was no linkage involved. That is the penalty for "not quite measuring the right construct." So, these linked values should only be used in studies in which the research design offers no other choice. When linked scores are necessary, we recommend the use of the average parameters in Table 5.

We observe explicitly that projected linkage between two measures is not symmetrical: The adult score projected from a pediatric score does not project back to the same pediatric score. Because the relation between the two scales is not perfect, there is some regression to the mean in both projections; this is required to minimize the error of prediction. This is a limitation of any linkage between scales that do not measure the same constructs.

This study was limited to those that could read and respond in English; thus, we cannot generalize the linking metric to other language versions of PROMIS. This study used quantitative methods to link the pediatric and adult PROMIS measures; however, further investigation is needed using qualitative methods to determine if adolescents and young adults conceptualize these symptoms in the same way. Generalizability of these algorithms to older individuals (e.g., >24 years) or younger children (i.e., <14) will need to be explored. However, a key strength of this study included the inclusion of adolescents and young adults with diverse health conditions, disabilities, and diseases.

Conclusion

This study used a relatively new linking method, calibrated projection, to create a measurement system that extends the use of PROMIS measures from age 8 years into adulthood. Using the regression coefficients and mean square error estimates in Table 5, one can estimate scores using the equations described in this paper that would likely have been obtained on one PROMIS version to the other. For example, it is estimated that an individual who scored a 40 on the PROMIS pediatric Anxiety measure would have likely scored a (20.23 + 0.68*40 =) 47.43 on the adult measure. Access to these linking metrics will facilitate research when both children and adults are included in the study or when a study follows a child longitudinally into adulthood.

Acknowledgments

PROMIS® was funded with cooperative agreements from the National Institutes of Health (NIH) Common Fund Initiative (Northwestern University, PI: David Cella, PhD, U54AR057951, U01AR052177; Northwestern University, PI: Richard C. Gershon, PhD, U54AR057943; American Institutes for Research, PI: Susan (San) D. Keller, PhD, U54AR057926; State University of New York, Stony Brook, PIs: Joan E. Broderick, PhD and Arthur A. Stone, PhD, U01AR057948, U01AR052170; University of Washington, Seattle, PIs: Heidi M. Crane, MD, MPH, Paul K. Crane, MD, MPH, and Donald L. Patrick, PhD, U01AR057954; University of Washington, Seattle, PI: Dagmar Amtmann, PhD, U01AR052171; University of North Carolina, Chapel Hill, PI: Harry A. Guess, MD, PhD (deceased), Darren A. DeWalt, MD, MPH, U01AR052181; Children's Hospital of Philadelphia, PI: Christopher B. Forrest, MD, PhD, U01AR057956; Stanford University, PI: James F. Fries, MD, U01AR052158; Boston University, PIs: Alan Jette, PT, PhD, Stephen M. Haley, PhD (deceased), and David Scott Tulsky, PhD (University of Michigan, Ann Arbor), U01AR057929; University of California, Los Angeles, PIs: Dinesh Khanna, MD (University of Michigan, Ann Arbor) and Brennan Spiegel, MD, MSHS, U01AR057936; University of Pittsburgh, PI: Paul A. Pilkonis, PhD, U01AR052155; Georgetown University, PIs: Carol. M. Moinpour, PhD (Fred Hutchinson Cancer Research Center, Seattle) and Arnold L. Potosky, PhD, U01AR057971; Children's Hospital Medical Center, Cincinnati, PI: Esi M. Morgan DeWitt, MD, MSCE, U01AR057940; University of Maryland, Baltimore, PI: Lisa M. Shulman, MD, U01AR057967; and Duke University, PI: Kevin P. Weinfurt, PhD, U01AR052186). NIH Science Officers on this project have included Deborah Ader, PhD, Vanessa Ameen, MD (deceased), Susan Czajkowski, PhD, Basil Eldadah, MD, PhD, Lawrence Fine, MD, DrPH, Lawrence Fox, MD, PhD, Lynne Haverkos, MD, MPH, Thomas Hilton, PhD, Laura Lee Johnson, PhD, Michael Kozak, PhD, Peter Lyster, PhD, Donald Mattison, MD, Claudia Moy, PhD, Louis Quatrano, PhD, Bryce Reeve, PhD, William Riley, PhD, Peter Scheidt, MD, Ashley Wilder Smith, PhD, MPH, Susana Serrate-Sztein, MD, William Phillip Tonkins, DrPH, Ellen Werner, PhD, Tisha Wiley, PhD, and James Witter, MD, PhD. The contents of this article uses data developed under PROMIS. These contents do not necessarily represent an endorsement by the US Federal Government or PROMIS. See www.nihpromis.org for additional information on the PROMIS® initiative.

References

1. DeWalt DA, Rothrock N, Yount S, Stone AA. Evaluation of item candidates: The PROMIS qualitative item review. Medical Care. 2007; 45(5 Suppl 1):S12–21. 00005650-200705001-00003 [pii]. 10.1097/01.mlr.0000254567.79743.e2 [PubMed: 17443114]
2. Irwin DE, Varni JW, Yeatts K, DeWalt DA. Cognitive interviewing methodology in the development of a pediatric item bank: A Patient Reported Outcomes Measurement Information System (PROMIS) study. Health and Quality of Life Outcomes. 2009; 7:3. 1477-7525-7-3 [pii]. 10.1186/1477-7525-7-3 [PubMed: 19166601]
3. Dorans NJ. Linking scores from multiple health outcome instruments. Qual Life Res. 2007; 16(Suppl 1):85–94.10.1007/s11136-006-9155-3 [PubMed: 17286198]
4. Bjorner JB, Chang CH, Thissen D, Reeve BB. Developing tailored instruments: item banking and computerized adaptive assessment. Qual Life Res. 2007; 16(Suppl 1):95–108.10.1007/s11136-007-9168-6 [PubMed: 17530450]
5. Thissen D, Reeve BB, Bjorner JB, Chang CH. Methodological issues for building item banks and computerized adaptive scales. Qual Life Res. 2007; 16(Suppl 1):109–119.10.1007/s11136-007-9169-5 [PubMed: 17294284]
6. McHorney CA, Cohen AS. Equating health status measures with item response theory: illustrations with functional status items. Med Care. 2000; 38(9 Suppl):II43–59.10.1097/00005650-200009002-00008 [PubMed: 10982089]
7. Haley SM, Ni P, Lai JS, Tian F, Coster WJ, Jette AM, Straub D, Cella D. Linking the activity measure for post acute care and the quality of life outcomes in neurological disorders. Arch Phys Med Rehabil. 2011; 92(10 Suppl):S37–43.10.1016/j.apmr.2011.01.026 [PubMed: 21958921]
8. Masse LC, Allen D, Wilson M, Williams G. Introducing equating methodologies to compare test scores from two different self-regulation scales. Health Educ Res. 2006; 21(Suppl 1):i110–120.10.1093/her/cyl088 [PubMed: 16926231]
9. Orlando M, Sherbourne CD, Thissen D. Summed-score linking using item response theory: application to depression measurement. Psychol Assess. 2000; 12(3):354–359. [PubMed: 11021160]
10. Tian F, Ni P, Mulcahey MJ, Hambleton RK, Tulsky D, Haley SM, Jette AM. Tracking functional status across the spinal cord injury lifespan: linking pediatric and adult patient-reported outcome scores. Arch Phys Med Rehabil. 2014; 95(11):2078–2085. e2015.10.1016/j.apmr.2014.05.023 [PubMed: 24933214]
11. Fisher WP Jr, Eubanks RL, Marier RL. Equating the MOS SF36 and the LSU HSI Physical Functioning Scales. J Outcome Meas. 1997; 1(4):329–362. [PubMed: 9661727]
12. Bond, TG.; Fox, CM. Applying the Rasch model: Fundamental measurement in the human sciences. Mahwah, NJ: Lawrence Erlbaum Associates; 2007.
13. Holzner B, Bode RK, Hahn EA, Cella D, Kopp M, Sperner-Unterweger B, Kemmler G. Equating EORTC QLQ-C30 and FACT-G scores and its use in oncological research. Eur J Cancer. 2006; 42(18):3169–3177.10.1016/j.ejca.2006.08.016 [PubMed: 17045472]
14. Wahl I, Lowe B, Bjorner JB, Fischer F, Langs G, Voderholzer U, Aita SA, Bergemann N, Brahler E, Rose M. Standardization of depression measurement: a common metric was developed for 11

self-report depression measures. J Clin Epidemiol. 2014; 67(1):73–86.10.1016/j.jclinepi.2013.04.019 [PubMed: 24262771]

15. Huang IC, Wu AW, Frangakis C. Do the SF-36 and WHOQOL-BREF measure the same constructs? Quality of Life Research. 2006; 15(1):15–24.10.1007/s11136-005-8486-9 [PubMed: 16411027]
16. Thissen, D.; Liu, Y.; Magnus, B.; Quinn, H. Extending the use of multidimensional IRT calibration as projection: Many-to-one linking and linear computation of projected scores. In: van der Ark, LA.; Bolt, DM.; Show, SM.; Douglas, JA.; Wang, WC., editors. New Methods and Applications in Psychometrics: The 79th Annual Meeting of the Psychometric Society. New York: Springer; (in press)
17. Irwin DE, Stucky B, Langer MM, Thissen D, Dewitt EM, Lai JS, Varni JW, Yeatts K, DeWalt DA. An item response analysis of the pediatric PROMIS anxiety and depressive symptoms scales. Quality of Life Research. 2010; 19(4):595–607.10.1007/s11136-010-9619-3 [PubMed: 20213516]
18. Irwin DE, Stucky BD, Langer MM, Thissen D, Dewitt EM, Lai JS, Yeatts KB, Varni JW, Dewalt DA. PROMIS Pediatric Anger Scale: An item response theory analysis. Quality of Life Research. 2012; 21(4):697–706.10.1007/s11136-011-9969-5 [PubMed: 21785833]
19. Pilkonis PA, Choi SW, Reise SP, Stover AM, Riley WT, Cella D. Item banks for measuring emotional distress from the Patient-Reported Outcomes Measurement Information System (PROMIS(R)): depression, anxiety, and anger. Assessment. 2011; 18(3):263–283.10.1177/1073191111411667 [PubMed: 21697139]
20. McPherson M, Arango P, Fox H, Lauver C, McManus M, Newacheck PW, Perrin JM, Shonkoff JP, Strickland B. A new definition of children with special health care needs. Pediatrics. 1998; 102(1 Pt 1):137–140.10.1542/peds.102.1.137 [PubMed: 9714637]
21. Neff JM, Sharp VL, Muldoon J, Graham J, Popalisky J, Gay JC. Identifying and classifying children with chronic conditions using administrative data with the clinical risk group classification system. Ambul Pediatr. 2002; 2(1):71–79.10.1367/1539-4409(2002)002<0071:iaccwc>2.0.co;2 [PubMed: 11888441]
22. Bethell CD, Read D, Stein RE, Blumberg SJ, Wells N, Newacheck PW. Identifying children with special health care needs: development and evaluation of a short screening instrument. Ambul Pediatr. 2002; 2(1):38–48.10.1367/1539-4409(2002)002<0038:icwshc>2.0.co;2 [PubMed: 11888437]
23. Gershon R, Rothrock NE, Hanrahan RT, Jansky LJ, Harniss M, Riley W. The development of a clinical outcomes survey research application: Assessment Center. Qual Life Res. 2010; 19(5): 677–685.10.1007/s11136-010-9634-4 [PubMed: 20306332]
24. Irwin DE, Stucky BD, Thissen D, Dewitt EM, Lai JS, Yeatts K, Varni JW, DeWalt DA. Sampling plan and patient characteristics of the PROMIS pediatrics large-scale survey. Quality of Life Research. 2010; 19(4):585–594.10.1007/s11136-010-9618-4 [PubMed: 20204706]
25. Dorans NJ. Equating, concordance, and expectation. Applied psychological measurement. 2004; 28:227–246.10.1177/0146621604265031
26. Dorans, NJ.; Walker, ME. Sizing up linkages. In: Dorans, NJ.; Pommerich, M.; Hollands, PW., editors. Linking and aligning scores and scales. New York: Springer; 2007. p. 179-198.
27. Dorans NJ, Holland PW. Population Invariance and the Equatability of Tests: Basic Theory and the Linear Case. Journal of Educational Measurement. 2000; 37(4):281–306.10.1002/j.2333-8504.2000.tb01842.x
28. Cai, L.; Thissen, D.; du Toit, SHC. IRTPRO for Windows. Lincolnwood, IL: Scientific Software International; 2011.
29. Samejima, F. Graded response model. In: van der Liden, WJ.; Hambleton, RK., editors. Handbook of Modern Item Response Theory. NY: Springer; 1997. p. 85-100.
30. Samejima F. Estimation of Latent Ability Using a Response Pattern of Graded Scores. Psychometrika monograph supplement. 1969
31. Thissen D, Varni JW, Stucky BD, Liu Y, Irwin DE, Dewalt DA. Using the PedsQL 3.0 asthma module to obtain scores comparable with those of the PROMIS pediatric asthma impact scale (PAIS). Qual Life Res. 2011; 20(9):1497–1505.10.1007/s11136-011-9874-y [PubMed: 21384264]

32. Kolen, MJ.; Brennan, RL. Test equating, scaling, and linking. Methods and Practice. 2. New York: Springer; 2004.

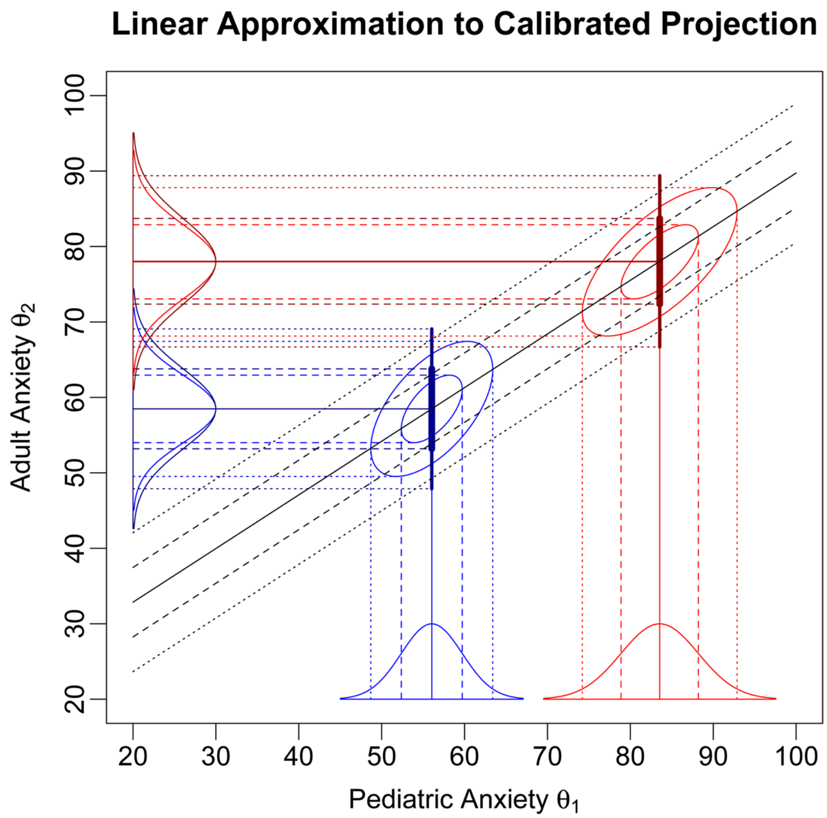

Figure 1.

The *x*-axis variable is θ_1, the underlying construct measured by the first scale and the *y*-axis variable is θ_2, the underlying construct measured by the second scale; the illustration is for the PROMIS pediatric and adult Anxiety scales. Both scales report scores in *T*-score units. The IRT model distributions on θ_1 for two response patterns on the pediatric Anxiety scale are shown along the *x*-axis, along with the corresponding bivariate distributions, and those implied by calibrated projection on θ_2. The darker blue and red lines and curves on the θ_2 axis show the normal approximations to the projected posterior distributions computing using the linear approximation, based on the regression (solid black) line, and the combination of the variances of the θ_1 distributions and the variance around the regression line.

Author Manuscript

Author Manuscript

Author Manuscript

Table 1

PROMIS Pediatric and Adult measures by domain[a]

Domain	Pediatric form(s)	Adult form
Physical Function	PROMIS Ped SF v1.0 – Mobility 8a PROMIS Ped SF v1.0 – Upper Extremity 8a	PROMIS SF v1.0 – Physical Function 10a
Pain	PROMIS Ped SF v1.0 – Pain Interference 8a	PROMIS SF v1.0 – Pain Interference 8a
Fatigue	PROMIS Ped SF v1.0 – Fatigue 10a	PROMIS SF v1.0 – Fatigue 8a
Social Health	PROMIS Ped SF v1.0 – Peer Relations 8a	PROMIS SF v2.0 – Social Health: Emotional Support
Depression	PROMIS Ped SF v1.0 – Depressive Symptoms 8a	PROMIS SF v1.0 – Depression 8b
Anxiety	PROMIS Ped SF v1.0 – Anxiety 8a	PROMIS SF v1.0 – Anxiety 8b
Anger	PROMIS Ped SF v1.0 – Anger 6a	PROMIS SF v1.0 – Anger 8a

[a]Note: The number listed after the domain name indicates the number of items in the short form.

Table 2

Demographic and clinical characteristics of the two samples.

	Special Health Care Needs		Disabilities	
	Adolescents (14–17 years) N = 415	**Young Adults (18–20 years) N = 459**	**Adolescents (14–17 years) N =188**	**Young Adults (18–24 years) N =453**
Age (mean, (SD))	15.63 (1.20)	18.93 (.75)	15.6 (1.26)	21.43 (2.21)
Sex				
Male	214 (51.6%)	194 (42.3%)	112 (59.6%)	291 (64.2%)
Female	200 (48.2%)	265 (57.7%)	75 (39.9%)	161 (35.5%)
Missing	1 (0.2%)	0	1 (0.5%)	1 (0.2%)
Ethnicity				
Hispanic	138 (33.3%)	196 (42.7%)	13 (6.9%)	40 (8.8%)
Non-Hispanic	277 (66.7%)	263 (57.3%)	153 (81.4%)	366 (80.8%)
Not provided	0	0	22 (11.7%)	47 (10.4%)
Race				
White	216 (52.0%)	228 (49.7%)	150 (79.8%)	366 (80.8%)
Black or African Am	93 (22.4%)	89 (19.4%)	20 (10.6%)	44 (9.7%)
Asian	34 (8.2%)	40 (8.7%)	5 (2.7%)	9 (2.0%)
Other	51 (12.3%)	63 (13.7%)	11 (5.8%)	20 (4.4%)
Multiple Races	14 (3.4%)	19 (4.1%)	2 (1.0%)	4 (0.8%)
Missing	7 (1.7%)	20 (4.4%)	0	8 (1.7%)
Health Condition				
Hypertension	81 (19.5%)	118 (25.7%)	---	---
ADHD	138 (33.3%)	108 (23.5%)	---	---
Mental health	104 (25.1%)	95 (20.7%)	---	---
Kidney disease	17 (4.1%)	15 (3.3%)	---	---
Chronic pain	88 (21.2%)	103 (22.4%)	---	---
Asthma	93 (22.4%)	104 (22.7%)	---	---
Thyroid disease	14 (3.4%)	18 (3.9%)	---	---
Overweight	67 (16.1%)	85 (18.5%)	---	---
Rheumatic disease	12 (2.9%)	13 (2.8%)	---	---
Born prematurely	18 (4.3%)	20 (4.4%)	---	---
Blind	7 (1.7%)	6 (1.3%)	---	---
Deaf	10 (2.4%)	8 (1.7%)	---	---
Needs walking assist.	7 (1.7%)	8 (1.7%)	---	---
Cancer	8 (1.9%)	18 (3.9%)	---	---
Diabetes	37 (8.9%)	52 (11.3%)	---	---
Sickle cell disease	7 (1.7%)	8 (1.7%)	---	---
Intestinal disease	19 (4.6%)	17 (3.7%)	---	---
Heart disease	11 (2.7%)	6 (1.3%)	---	---

	Special Health Care Needs		Disabilities	
	Adolescents (14–17 years) N = 415	**Young Adults (18–20 years) N = 459**	**Adolescents (14–17 years) N =188**	**Young Adults (18–24 years) N =453**
Epilepsy	15 (3.6%)	17 (3.7%)	---	---
Allergies	98 (23.6%)	81 (17.6%)	---	---
Cerebral palsy	6 (1.4%)	4 (0.9%)	99 (52.7%)	98 (21.6%)
Spinal cord injury	---	---	53 (28.2%)	189 (41.7%)
Traumatic brain injury	---	---	36 (19.1%)	166 (36.6%)

Table 3

Descriptive statistics for the scores: The means and standard deviations (SD) of the scale scores (EAP[θ]) for the pediatric and adult scales, and the correlation coefficient (r) between the pediatric and adult scores.

		PROMIS Pediatric		PROMIS Adult		
Domain	**Sample**	**Mean**	**SD**	**Mean**	**SD**	***r***
Depressive Symptoms	SHCN[a]	56.8	12.0	56.7	11.1	0.86
	Disabilities[b]	48.4	10.4	48.6	9.2	0.84
Anxiety	SHCN	55.6	11.4	58.1	10.3	0.82
	Disabilities	46.6	10.9	50.7	9.1	0.77
Anger	SHCN	54.0	12.4	55.9	11.9	0.84
	Disabilities	46.7	10.9	49.4	10.0	0.80

[a]SHCN = adolescents and young adults with special health care needs

[b]Disabilities = adolescents and young adults with physical or cognitive disability

Table 4

Root Expected Mean Square Difference (REMSD) statistics for the sex difference, and the difference between adolescents and young adults, and the estimated latent-variable correlation ρ between the pediatric and adult constructs with its standard error (*s.e.*).

Domain	Sample	REMSD Sex	REMSD Age[a]	ρ	*s.e.* (ρ)
Depressive Symptoms	SHCN[b]	0.128	0.020	0.93	0.01
	Disabilities[c]	0.020	0.027	0.94	0.01
Anxiety	SHCN	0.003	0.007	0.89	0.01
	Disabilities	0.038	0.015	0.87	0.01
Anger	SHCN	0.004	0.052	0.91	0.01
	Disabilities	0.033	0.027	0.91	0.01

[a]Age categorized by adolescent (14–17 years) and young adult (18–24 years)

[b]SHCN = adolescents and young adults with special health care needs

[c]Disabilities = adolescents and young adults with a physical or cognitive disability

Table 5

Regression coefficients β_0 and β_1, and MSE, that are used in equations 1 and 2 to compute scale scores by calibrated projection from θ_1 to θ_2, on the *T*-score scale used for PROMIS instruments; 95% confidence intervals in parentheses below the estimates.

		Pediatric to adult			Adult to pediatric		
Domain	**Sample**	β_0	β_1	**MSE**	β_0	β_1	**MSE**
Depressive Symptoms	SHCN	10.65 (6.58 – 14.72)	0.81 (0.77 – 0.85)	19.4	−3.12 (−7.79 – 1.55)	1.06 (1.01 – 1.11)	25.4
	Disabilities	12.29 (8.52 – 16.06)	0.76 (0.69 – 0.83)	12.5	−8.38 (−12.60 – −4.16)	1.16 (1.08 – 1.24)	19.2
	Average	**11.47**	**0.78**	**16.0**	**−5.75**	**1.11**	**22.3**
Anxiety	SHCN	18.64 (16.41 – 20.87)	0.71 (0.67 – 0.75)	21.2	−9.48 (−13.23 – −5.73)	1.12 (1.06 – 1.18)	33.4
	Disabilities	21.83 (19.45 – 24.21)	0.64 (0.59 – 0.69)	20.5	−15.24 (−19.86 – −10.62)	1.19 (1.10 – 1.28)	38.2
	Average	**20.23**	**0.68**	**20.8**	**−12.36**	**1.16**	**35.8**
Anger	SHCN	14.00 (11.74 – 16.26)	0.78 (0.74 – 0.82)	27.8	−5.98 (−9.16 – −2.80)	1.07 (1.02 – 1.12)	38.4
	Disabilities	17.83 (15.52 – 20.14)	0.70 (0.65 – 0.75)	20.5	−13.35 (−17.47 – −9.23)	1.19 (1.11 – 1.27)	35.0
	Average	**15.91**	**0.74**	**24.2**	**−9.67**	**1.13**	**36.7**

Table 6

Proportions of values of values of EAP[θ_2] for each scale combination that are within ±1 SD and ±2 SD of the values obtained using the linear approximation to calibrated projection.

	Special Health Care Needs (SHCN) Sample				Disability Sample			
	Pediatric to adult		Adult to pediatric		Pediatric to adult		Adult to pediatric	
	±1 SD	±2 SD	±1 SD	±2 SD	±1 SD	±2 SD	±1 SD	±2 SD
Domain	SHCN parameters				Disability parameters			
Depressive Symptoms	0.70	0.93	0.71	0.93	0.70	0.93	0.70	0.93
Anxiety	0.69	0.93	0.71	0.92	0.70	0.92	0.71	0.95
Anger	0.71	0.95	0.74	0.94	0.65	0.95	0.73	0.95
	Disability parameters				SHCN parameters			
Depressive Symptoms	0.58	0.88	0.65	0.89	0.76	0.95	0.73	0.95
Anxiety	0.65	0.92	0.71	0.93	0.70	0.94	0.69	0.95
Anger	0.62	0.91	0.70	0.93	0.77	0.96	0.76	0.95
	Average parameters				**Average parameters**			
Depressive Symptoms	0.75	0.92	0.68	0.92	0.74	0.94	0.72	0.94
Anxiety	0.67	0.92	0.72	0.93	0.70	0.94	0.72	0.95
Anger	0.69	0.93	0.72	0.93	0.76	0.96	0.74	0.95